MATI-218US

Appln. No.: 10/524,965
Amendment Dated August 28, 2006
Reply to Office Action of June 13, 2006

Amendment to the Abstract:

A new Abstract is provided. A clean version is attached on a separate piece of paper.

ABSTRACT

A method for generating structured documents files from a document image includes segmenting the document image into one or zones, and converting the respective text images within the one or more zones to digital text. One or more zones includes a respective text image. The method further includes automatically identifying layout information for each of the one or more zones, labeling each of the one or more zones in accordance with a schema, and automatically associating marked up language tags with the labeled zones to generate the structured document files responsive to the identified layout information a model file.

Attachment